

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 13:18:50 ; Search time 41 seconds
(without alignments)
868.274 Million cell updates/sec

Title: US-09-494-297-2
Perfect score: 3945
Sequence: 1 MKKTRFPNKNITNTQRLVLS.....IAGISLGIMGIIHTRIRKHD 757

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	148	3.8	1183	1 CNA_STPAU	Q53654 staphylococ
2	139	3.5	621	1 HTPG_HELPY	P56116 heliobacte
3	139	3.5	1433	1 SUBF_BACSU	P16397 bacillus su
4	139	3.5	2366	1 TOXB_CLODI	P18177 clostridium
5	137	3.5	1177	1 V307_MYCGE	P47549 mycoplasma
6	136.5	3.5	4590	1 FATH_HUMAN	Q14517 homo sapien
7	135	3.4	1176	1 SLAP_BACSH	P38537 bacillus sp
8	135	3.4	1228	1 SLAP_BACST	P35825 bacillus st
9	135	3.4	1301	1 SAC3_YEAST	P46674 saccharomyc
10	132.5	3.4	1116	1 SLPH_BRECH	P38538 brevbacill
11	132.5	3.4	3097	1 CADN_DROME	O15943 drosophila
12	130.5	3.3	881	1 LHS1_YEAST	P36018 saccharomyc
13	130.5	3.3	908	1 DPO1_BORBU	O51498 borrelia bu
14	130.5	3.3	1020	1 RPRG_HUMAN	O32834 homo sapien
15	130	3.3	621	1 HTPG_HELPY	O92mm2 helicobacte
16	130	3.3	1372	1 FUSL_SCHPO	O10719 schizosacch
17	130	3.3	3381	1 PGCY_BOVIN	P81282 bos taurus
18	129	3.3	461	1 P55G_HUMAN	Q32569 homo sapien
19	129	3.3	1292	1 RPOC_MYCGE	P47582 mycoplasma
20	128.5	3.3	914	1 PBPB_BACSU	P39793 bacillus su
21	128.5	3.3	1182	1 RPOC_CLOAB	O97eh0 clostridium
22	128.5	3.3	1828	1 MAP2_MOUSE	P20357 mus musculu
23	128	3.2	1394	1 HAP_HAERIN	P45387 hemophilus
24	127	3.2	1628	1 NAGH_CLOPE	P26831 clostridium
25	127	3.2	1650	1 VIT6_CAEEL	P18948 caenorhabdi
26	126.5	3.2	822	1 DEXT_STREL	O59978 streptococc
27	126.5	3.2	1772	1 MSPI_PLAYO	P13828 plasmodium
28	126	3.2	1036	1 Y414_MYCGE	P47653 mycoplasma
29	126	3.2	1409	1 HAPL_HAERIN	P44596 hemophilus
30	125	3.2	700	1 HSPC_DICDI	P54651 dictyosteli
31	125	3.2	862	1 SLA2_BACAN	P94217 bacillus an
32	124.5	3.2	461	1 P55G_BOVIN	O46404 bos taurus
33	124.5	3.2	1251	1 RBP2_PLAVB	Q00799 plasmodium

34	124.5	3.2	3060	1 BPER_HUMAN	O84x8 homo sapien
35	124.5	3.2	6359	1 BACC_BACLI	O68008 b bacitraci
36	124	3.1	2334	1 WAPB_BACSU	O07833 bacillus su
37	123.5	3.1	1053	1 SLPM_BACBR	P06546 bacillus br
38	123.5	3.1	1358	1 SIR4_YEAST	P11978 saccharomyc
39	123.5	3.1	3063	1 CALC_HUMAN	O99715 homo sapien
40	123	3.1	461	1 P55G_MOUSE	O64143 mus musculu
41	123	3.1	831	1 ION_HELPY	O97113 helicobacte
42	123	3.1	980	1 BOB1_YEAST	P38041 saccharomyc
43	123	3.1	1022	1 SCA4_RICPR	O92d49 rickettsia
44	123	3.1	1055	1 CC28_SCHPO	O10752 schizosacch
45	122.5	3.1	681	1 BRC2_HAERO	O01409 halocynthia

ALIGNMENTS

RESULT 1

CNA_STPAU STANDARD: PRT: 1183 AA.
ID Q53654;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RT Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
RT staphylococcus aureus collagen adhesin."
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeoek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from staphylococcus aureus."
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Smeers J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin."
RL Nat. Struct. Biol. 4:833-838(1997).
CC -I- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -I- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL: M61736; AAA20874.1; -.
DR PDB: 1AMX; 24-JUN-98.
DR PDB: 1D20; 27-SEP-00.
DR PDB: 1D2P; 27-SEP-00.
DR InterPro: IPR001899; Gram_pos_anchor.
DR TIGRGRAMS: TIGR01167; LPXTG_anchor; 1.
DR PROSITE: PS50847; GRAM_POS_ANCHORING; FALSE_NEG.
DR Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
KW SIGNAL.
FT CHAIN 1 29
FT PROPEP 30 1154
FT DOMAIN 151 1183
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
FT SITE 1151 1155
FT MOD_RES 1154 1154
FT STRAND 174 179
FT TURN 182 183
FT STRAND 187 194
FT TURN 196 197
FT STRAND 201 201
FT STRAND 205 211
FT STRAND 215 228
FT TURN 229 230
FT STRAND 232 234
FT HELIX 239 246
FT TURN 248 249
FT STRAND 251 255
FT TURN 256 259
FT STRAND 260 265
FT HELIX 267 270
FT TURN 271 272
FT STRAND 273 283
FT TURN 286 287
FT STRAND 290 299
FT STRAND 301 301
FT TURN 302 303
FT STRAND 307 311
FT STRAND 314 317
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match      3.8%; Score 148; DB 1; Length 1183;
Best Local Similarity 19.7%; Pred. NO. 1;
Matches 165; Conservative 123; Mismatches 299; Indels 252; Gaps 44;

OY 64 PSSSEYRWY---GYESYVGRHPYK-----QFVAVHDLRYNLEGSRSY---QVYCF 109
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 PEDTTHVWFELNINNEKSYSKDITKIDGGQGLDLSPLNINVTGTHSNYSGSAIT 241
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 110 NLKKAFFLS-----DSSVAKWKTKHDGISTKPEDYAMSPRTIG----- 148
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 242 DEKAPF--OSKITVDNTKNTIDVTIPQGYGSYFSINXK-----TKTINEQKEFEVNS 295
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 149 -----DELNOKLRAVYNGHPONANGIMG-----LEPLNAIRVTOEAVWYSDNA 194
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 296 QAWYQHGKEEVNGK--SENHVYHNANAGISGTIVKGLKYLKQKQDKT-----A 344
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 195 PISNPDESPK--RESESNLVSTQSLMRQALQDLDPNLATKPKQVDPDFOLSIPESE- 252
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 345 PLAN--VRFKLSKSDSVKNDQ-----KEIEITDANGIANIKALPSGDYILKEIAPR 397
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 253 ----DKGDKY-----NKGYONLSGGL-----VPTKPTPEDPPMPNPQPTTSY 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 398 PYTFDDEKPEYPTMKDPTDNOGYFTTLENKAKLEKTKDVSQAOKWEGSTOKYKP-----TI 451
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 294 LIRKVAIGYS-----KLLGATLQTLGDNVNSQARVFSSND--IGERIELSDGI 342
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 452 YRKLKYKODDNQNTPTPDKAIEIKLEEDGTTKVTWSNL-----PENDKNKAIRK----- 498

```

```

OY 343 YLTLELN-----SPAGYSIAE-----PI-TRKYBAGKYVYIIIDGKOIENPKEL- 385
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 499 YLVEVNAQGEDTTPREGYTKKENGVLVNTNERKPLETTSISGEKWDKDNQDKRREK-- 556
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 386 VEPYSVAVNDPEFSPVLTQNTAKRY-----AKKNKSSQVYVCENADLKSPSEDC 440
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 557 --VSYNLLANGERKVTLDVTSFTNMKYEFKDLPKYDEG--KLEIYVTEBDHVADYTTDLN 612
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 441 GKTMPTDEFTGGEKYTHIAGRDLEFKYTVKRPDRDPPTFLKHKKVLEK----- 488
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 613 GTITTKYTTGELTSAYTKWMDNNODGKRPFIEIKVELYQDSKATGKTAIINESNNWTH 672
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 489 -----GYREKQALIEYSGLETQLRATQ-----LAIYFTDSALDKRLKYD 532
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 673 TWTGLDEKAKGQGVKVTVEELTVKGYTTHVDNNDGNLVTNKKYPTFTSISGEKWD- 721
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 533 HGFQDNDSTLAVAKILVEAODSNPQLTDLDFIPINNKKYQSL--IGTQHPDEDVD 589
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 732 ----DKDN-----QDGKREKVSAN-LIADGERKVTLDVTSFTNMKYE-FKD 772
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 590 IIRMEDEKREYIPYTHNLTKRYTGLAGDRTKDFHFEIE-----LKNKQELLSQYVKT- 643
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 773 LPKYDESKTI-----EYIVT--EDHVKYTTDINGTITTKYTPGETSAYTKW 820
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 644 -DKTNLE-----FKDGRAT-----INLKHESLTLAGIPE-----GYSYLKE- 680
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 821 DDNNQDGKRPFIEIKVELYQDSKATGKTAIINESNNWTHWTGLDEKAKGQVYVLEL 880
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 681 TDESGYKVKVNSQEVANATVSK-----IGITSEDLAPENKKEPVVPGVDOKT--NG 731
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 881 TKVKGYYTHVDNNDGNLVTNKKYPTFTSISGEKWDKDNQDKRPERKVSYNLLANG 939
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
HTPG_HELPY
ID HTPG_HELPY STANDARD; PRT; 621 AA.
AC P56116;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperone protein htpg (Heat shock protein htpg) (High temperature
DE protein G).
GN HTPG OR HP0210.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteriaceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RC MEDLINE=97394467; PubMed=9252185;
RX Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).

DR EMBL: AE000541; AAD07278.1; -
 DR PIR: B64546; B64546.
 DR HSSP: P07900; 1YER.
 DR TIGR: HP0210.
 DR HAMAP: MF_00505; -; 1.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR001404; Hsp90.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF00183; HSP90-1.
 DR PRINTS: PR00775; HEATSHOCK90.
 DR SMART: SM00387; HATPase_c; 1.
 DR PROSITE: PS00298; HSP90; 1.
 KW Chaperone; ATP-binding; Heat shock; Complete proteome.
 FT DOMAIN 1 341 A; SUBSTRATE-BINDING (BY SIMILARITY).
 FT DOMAIN 342 547 B (BY SIMILARITY).
 FT DOMAIN 548 621 C; 28F198CIDC7EAB9A CRC64.
 SO SEQUENCE 621 AA; 71274 MW; 28F198CIDC7EAB9A CRC64.

Query Match 3.5%; Score 139; DB 1; Length 621;
 Best Local Similarity 21.9%; Pred. No. 1.2;
 Matches 130; Conservative 85; Mismatches 227; Indels 152; Gaps 30.

OY 201 ESFRRESSE-----LVSTSQLSLMRQALKQLDPLNLTAKMPQVDPDFQSLFSEEDKG- 255
 DB 28 EIFRELVSNSDALDKLNYLMLTDEKLT-----GLNTPPSIHLFSQCKTLTFTKNGI 82
 OY 256 --DKVN-----KGYONLSGGLVPTKPPRPDPMPNPQGTSTVLRKRYAIDY 303
 DB 83 GMDKNDLLEHGTAKSTKFNLSA-----LSGD-----KKKSALIGCGVGFY 127
 OY 304 SKLEGATLQLTGDNVNSFOARVSSN-----DIGERIELDGYTTLETNLSPAY----- 354
 DB 128 SAFVASKIVQTKRVNSDAQYAWMSDGKFEISECVKDSQGFTEITFLKDEDSHFASR 187
 OY 355 -----SIAPETFKVEAKVYTIIDGK-QIENPKNEIPEYVEAYNDFEER----- 400
 DB 188 WEIDSVKKYSEHLPFPI-----FLTYTDTKHEGEGDNCKEIKKEKCEQINOASALMKMK 243
 OY 401 SVLTTONYAKFYVAKNKGSSOVYCFNADLKSPDSDGCKTMDPTTGKVKYTHI-- 458
 DB 244 SELDKDKYEFYQSAHNSPEPLSTIHK-----VEGSLTYTTTFY 284
 OY 459 ---AGRDLFK-----YTVKRPDTPD-----TFLKHKKVLEKG-----YREKG 494
 DB 285 IPSTAPEDMFVRDVKSGVKLYKRVFTDDEKELLPSYLRVKGVIDSEDLPLVNSRE-- 342
 OY 495 QAIYESGLTQLRAATQALATYFTDSAELEKDKAKDYGCDMDSTLANAKLIVE--Y 552
 DB 343 --IIQQNKIIANISASVKI--LSEIERLSKDE-KNYKFEY-----PRGVLKEGLY 391
 OY 553 AQDSNPOLTDLDFEFPNNKKYOSLIGTOWHPEDLVDIRBDEKKEVIPV--HNLTLEKT 611
 DB 392 GDFENKEKLELELRFYSKD--EKLSLKEKLENL-----KENQSTIYLLGENDLKA 444
 OY 612 VTGLAGRTKDFHEIELKNKOE--LLSQTVKTDKINLEFKDGKATINLKHGESLLOG 669
 DB 445 SPL--EKYAKGQGVALLSDIDAFAVMPGVNEYDKT--PRDA-----SHSLSLKELG 494
 OY 670 LPE-----GYSLVKTEDSEGYKVKVNSQEVANATYSKTIGTSDTLAEENN 716
 DB 495 LEEIHDEVKQFKDLMAFE--ENLKEIKGVELSHLSAVALIGDEONAMMAN 547

RESULT 3
 SUBF_BACSU STANDARD: PRT: 1433 AA.
 AC P16397;
 DT 01-AUG-1990 (Rel. 15, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (Rp-I protease)
 DE (90 kDa serine proteinase).
 GN BPR OR BPF.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_Taxid=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
 RX MEDLINE=90216713; PubMed=2106512;
 RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Thierault K.A.,
 RA Pero J.;
 RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
 RT and cloning of the gene";
 RL J. Bacteriol. 172:1470-1477(1990).
 RN [2]
 RP REVISIONS.
 RC STRAIN=168;
 RX MEDLINE=90368623; PubMed=2118514;
 RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Thierault K.A.,
 RA Pero J.;
 RL J. Bacteriol. 172:5520-5521(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90216713; PubMed=2108961;
 RA Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
 RT "Cloning, genetic organization, and characterization of a structural
 RT gene encoding bacillopeptidase F from Bacillus subtilis";
 RL J. Biol. Chem. 265:6845-6850(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brigneau S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entlan K.D., Errington J., Fabel C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Guim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Jorhis B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolet C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowka A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaroni A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Welleneger T.,
 RA Winters P., Wipat A., Yamamoto H., Yaman K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis";
 RL Nature 390:249-256(1997).
 RN [5]
 RP SEQUENCE OF 1-211 FROM N.A.
 RX MEDLINE=89008108; PubMed=3139638;
 RA Beall B., Lowe M., Lutkenhaus J.;
 RT "Cloning and characterization of Bacillus subtilis homologs of
 RT Escherichia coli cell division genes ftsZ and ftsA";
 RL J. Bacteriol. 170:4855-4864(1988).
 RN [6]

RP SEQUENCE OF 1410-1433 FROM N.A.
 RC STRAIN-168 / Marburg;
 RX MEDLINE=50174995; PubMed=2106671;
 RA Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
 RT "Nucleotide sequence of the sporulation gene spoIIIGA from *Bacillus*
 RT *subtilis*.";
 RL Nucleic Acids Res. 18:657-657(1990).
 RP SEQUENCE OF 195-219.
 RC STRAIN-Natto 16;
 RA Kato T., Yamagata Y., Arai T., Ichishima E.;
 RT "Purification of a new extracellular 90-kDa serine proteinase with
 RT isoelectric point of 3.9 from *Bacillus subtilis* (natto) and
 RT elucidation of its distinct mode of action.";
 RL Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to peptidase family S8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M29035; AAA62679.1; -
 CC EMBL: J05400; AAA83362.1; -
 CC EMBL: Z99111; CAB13403.1; -
 CC EMBL: Z99112; CAB13404.1; -
 CC EMBL: M29630; AAA23458.1; -
 CC EMBL: X17344; CAA35224.1; -
 CC PIR: A36734; A36734.
 CC HSP: P00782; 2SMT.
 CC MEROPS: S08.017; -
 CC Subtilist: BG10233; hpr.
 CC Interpro: IPR000209; Peptidase_S8.
 CC Pfam: PF00082; Peptidase_S8; 1.
 CC PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
 CC PROSITE: PS00137; SUBTILASE_HIS; 1.
 CC PROSITE: PS00138; SUBTILASE_SER; 1.
 CC HydroLase: Serine protease; Zymogen; Signal; Complete proteome.
 FT SIGNAL 1 30
 FT PROPEP 31 194
 FT CHAIN 195 755
 FT PROPEP 756 1433
 FT ACT_SITE 227 227
 FT ACT_SITE 274 274
 FT ACT_SITE 452 452
 FT CONFLICT 219 219
 FT CONFLICT 393 393
 FT CONFLICT 829 834
 FT CONFLICT 836 841
 FT CONFLICT 844 852
 FT CONFLICT 853 1433
 FT SEQUENCE 1433 AA; 154577 MW; 98DF6846897807C9 CRC64;
 Query Match 3.5%; Score 139; DB 1; Length 1433;
 Best Local Similarity 20.7%; Pred. No. 4.3;
 Matches 165; Conservative 101; Mismatches 273; Indels 258; Gaps 43;
 QY 64 PDSSEY-----RWYGE-----SYV-----RGHYTKQFRVH-----DLRVNLEGRS 103
 DB 701 PDSGSLFLDKSRWNHLEDDFDYGVVLPBGEKNWEDAGVNGKTSSWTEEDLISAYKG 760
 QY 104 YQVYCFNLKAPPLGSDSVK--WYKKHGDISTKFEVDYMSPRITGDEL----- 151
 DB 761 Q-----NIQMFVLDGDESTAKEGW-----IDVVLSDKSACKTVKKKGLGYEK 806
 QY 152 --NOKLRVYVNGHPONAN-----GIMEGLEPLNA-IRVQOEAVMYSDNA----- 194
 DB 807 SGKQKKRVNPKAKKASANTAVKHNKALIQVLPRLKAGVSVEETGKSTSDSTGQYTL 866

QY 195 -----PISNDESEFKRESNEVLS-----TSQISLMRQALKQ-----LIDPNL---ATK 235
 DB 867 KHKAGDVTLAAEAVGYSKQTKOVSLKTDQTAQNFTEEMKGTGLKGTIVKTKTGPEVYG 926
 QY 236 MPKQVPPDFQSLFESDEKDKDKTKNGYQNLISGLVPTKPTTGPMPMPNQPTSVLI 255
 DB 927 ASYVVEADAVERPAMTNDKGEYMLEAVE-----GAYTIKAAPR----- 965
 QY 296 RKVAIGDYSKLECATLQ-----TGDNVSFOARFESSNDIGERIE 337
 DB 966 --YSDSEFVELKGDVTKETALRPFVGPBEIAYDGTAEANASYA---AGNCAVAKMT 1020
 QY 338 LSDG-----TYTTELNSPAGYSIAEPITFEVKAQYV---TIIDKQIENPKE 384
 DB 1021 LADGKDKMGLTGLFREMDEFPDPG-----TEFKVE---YDADGKDKGA---PGK 1067
 QY 385 IVEPYSVEAVN-----DPEEPSVLTONYAKFY-----YAKKNSSQVYVC 426
 DB 1068 IAGFNAELALNGEMTKVDLSSKGINVDKDFYLYIQSKDPDPSGLAMDETQNSGRW 1127
 QY 427 FNADLK-SPPSEDDG-----KTPEDFTTGEVK-YTHIAGRDLFEYTVKPRDTP 475
 DB 1128 QYIDGKKQPDGKADGNMIMALVDYEAAYEILSPDKSYTN---KD--SVYK-GNASP 1181
 QY 476 DTFELK-----HIKKYIEKGYRE-KGAQIEYSGITETQLRAAQLAI 515
 DB 1182 GTTVHIYNGEKEAGETKAADGTFHAGIILINKENELTATASDNGTTD---ASSPIV 1237
 QY 516 YVFTDSAEELDKDKDKOVHGFQDMNDSLVAAKLVEAADSNPQLDLDFFIPNNKYQ 575
 DB 1238 TLDOEKPELTLDNPKD---GKTKNEKELTVKAVN---SDN---LKD---KVAVGKA 1283
 QY 576 SLIGTQWHPEDLVYDIIRMEDEKKEVIRPYTHNLTKTYTGLAGRTKDHFREI----- 627
 DB 1284 TVADGSYSARILLENGNEIKVIATDLAGKTKTKVYI-----DVNDRKVISGLIP 1335
 QY 628 -ELKNNKQELLISQTVTKDKTNLEPKGDKATI-----NLKHGESLTQ-----GLPEG 673
 DB 1336 GEDKNLK---AGESVKIAFSAEDLATFTIRPLTNARASVONATELPLREISPGRYEG 1392
 QY 674 YSYLVKRETDSEGYKVV 690
 DB 1393 YWTATSTIRKAKAKVEY 1409
 RESULT 4
 TOXB_CLODI STANDARD; PRT; 2366 AA.
 ID TOXB_CLODI
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Toxin B.
 GN TOXB OR TCDB.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1496;
 RN 11
 RN SEQUENCE FROM N.A.
 RC STRAIN-VPI 10463;
 RX MEDLINE=90326540; PubMed=2374729;
 RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
 RT "Nucleotide sequence of Clostridium difficile toxin B gene.";
 RL Nucleic Acids Res. 18:4004-4004(1990).
 RN 12
 RN SEQUENCE FROM N.A.
 RC STRAIN-VPI 10463;
 RA von Eichel-Streiber C.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN 13
 RC SEQUENCE OF 1271-2366 FROM N.A.
 RC STRAIN-VPI 10463;

FT DOMAIN 92 151 SLH 2.
 FT DOMAIN 152 210 SLH 3.
 SQ SEQUENCE 1176 AA; 125225 MW; 1A9A458EF43788C CRC64;

Query Match 3.48; Score 135; DB 1; Length 1176;
 Best Local Similarity 18.58; Pred. No. 5.4;
 Matches 135; Conservative 101; Mismatches 256; Indels 238; Gaps 31;

166 NANGIMEGLRPL-----NAIRYTOEA-----VWYSSNAPISNPDESFKR 205
 59 DANGNFPNPKTISRAREATIFTNALTELEAGDVNFKQVADANY-DATAIYENGTFEG 117
 206 ESSSNLYSTQSLMRQALQOLDPNLTATMPKQVPDDFQSLTFESDKD----- 256
 118 VSATEFAPNKOIT-RSEAAKILVD-----AFELGEGDLESEFADAST 158
 257 --KYNKGYNLSSGLVPTKPPPTGPPMPNPQPTSVLIRKAYAGDYSKLEGATLQ 314
 159 VKPMASYLEIAVANGYIKGSEANGKTNLNPAPTR---QDPAY-VFSRTLENDATP 213
 315 TGNVNSFOARFESSNDIGRIELSDGTTLTLELSPAGYSIAEPIFFKY-----EAGKV 369
 214 KVDKIEVVDAKTLN-----VTLSDGTRETVLEKRLPENKEVEYFKIKDVEYKAKVT 266
 370 YTIIDGKOIENPKKEIPEYPSVAYNDEEFVLTQNYA-----KYYAKNNGS 420
 267 YVTTATAYKVSATNLKEVYVEFDGTVDKETEDANAYALKSGKITKYSVLAADNKTAT 326
 421 SQVYVCFN---ADLKSPDESDGKTM---TPPTTGEVAYTHIAG-RDLFKYTVKRDY 473
 327 VTLTDLKNNKKAISINVKAGDKETNKNVETAVDNKIPEVTEYKSLGTRAVYTL 386
 474 DDPDTFLKIKKIVIEKGYREKGOAIEYSGLTETQLRAATOLAIYFTDSALDKDL---- 529
 387 EP-----VENLSTNTLTDGA--YFGNYYMGAGNKTVLTTPYSSALSVDGKLTVSG 438
 530 -KTYHGFQDMN-----DSTLAVAKLIVEYADSNPPLDIDFTIPNN 571
 439 AKDFAGFVSLNSTHEEFKVEVDKAPVTEATLEFVTLTFSED-----IDMDYKASN 492
 572 -----NKYOSLI-GNQ-----WHPEDLVDI-----IRMDK 597
 493 VYKSGDSKKEASEFERIADNKKTFYFKGSEKTLPTGKVDVYVEDIKDYDNKIAKTKV 552
 598 EVIP-----VTHNLTKRTVTGLAGDRPKDFHFEIELKNNKOELLSTQ 640
 553 TVPELIDQTRPEYKRYTALDEKTIKTYFTSKTVGESAIKTG--NYTKKDKD-----V 604
 641 VKTDKTNLFEKDGKA-----TINKHGESLT-LOGLEPGYSYLVKETSDEG 685
 605 VSVDKVTYVDSKDSVYIIDLYSKVSGENTITIKNVKADATKLNMTMLDYGKFTRSDEG 664
 686 -----YKRYN-----SOEYANATVSKTG- 704
 665 PDYEHVINDAKAKKVVLFKDKKMDAASLADYSNYLVKINDTLQTLSEDAVATLSVNDAT 724
 705 ---ITSDEL 711
 725 VVITTFATETI 734

RESULT 8

SLAP_BACST
 ID SLAP_BACST STANDARD; PRT; 1228 AA.
 AC P35825:
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE S-layer protein precursor (Surface layer protein).
 GN SBA.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Geobacillus.
 NX NCBI_TaxId=1422;

FN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=PV72;
 RX MEDLINE=94320770; PubMed=8045409;
 RA Kuen B., Sleytr U.B., Lubitz W.;
 RT "Sequence analysis of the sba gene encoding the 130-kDa
 surface-layer protein of Bacillus stearothermophilus strain PV72.";
 RL Gene 145:115-120(1994).
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
 CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.

CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X71092; CA50409.1; -
 DR EMBL: AX000218; CAB7069.1; -
 DR PIR: I40468; 140468.
 KW Signal; Cell wall; S-layer.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1228 S-LAYER PROTEIN.
 SQ SEQUENCE 1228 AA; 131076 MW; 9F1383AD810C0B0F CRC64;

Query Match 3.48; Score 135; DB 1; Length 1228;
 Best Local Similarity 19.34; Pred. No. 5.8;
 Matches 128; Conservative 88; Mismatches 230; Indels 216; Gaps 33;

206 ESSSNLYSTQSLMRQALQOLDPNLTATMPKQVPDDFQSLTFESR-DKGDYKNGYON 264
 551 EKESGVVASE-----LTYNADAKNVTLPV--ADLKENTTYQIKKIKGLKSDGITE- 600
 265 LLSGLVPTKPPPTGPPMPNPQPTSVLIRKAYAGDYSKLEGATLQTL----- 315
 601 -LGTVEKIEYKERTQDLTAPVIAVSKN--GD-----AGAKTTEAQEFTVKF 645
 316 GDNVNSFOARFESSNDIG-----ERIELSDGTTL----- 345
 646 SEMINFNATTVSGSTITYGQVAVYKAGANLSALTSDIIPASVEAVTGDDGYKRYKVA 705
 346 TELNSPAGYSI-----AEPIFFKEAGK-----YTTI-IDGKOIENPKKEIPEYSE 392
 706 NQLEKRGYKLVFVGKATAPVADANANTLATNTYTTTTEGQDVATP--VTKVEKGD 763
 393 AYNDPEEFVLTQNYAKFYAKNNGSSQVYCFNADLKSPDESDGKTMTPDETGE 452
 764 SLKDADAVTTL-----NVDAQKFTIQFSEBELKTSSGSLVGKVT----- 804
 453 VKYTHIAGRDLEFYKVPDTPDTFLKIKKIVIEKGYREKGOAIEYSGLTETQLRAAQ 512
 805 -----VEKLNNGMVADAGTIVSAVKTDANGKYT 835
 513 LAIYFTDSAEIDKO-KLK-----DYHGGQNDSTLAVAKLIVEY-----A 553
 836 AAVVYTLGLDNNKDKAKRLRYVDKSDTDGADAGAVIKEDLLIRYNSMRHTVASAKA 895
 554 QDSNP-----POLDLDFEIPNNKRYOSLICTQHHDPDLVDIIMEEDKREYIPTVHNL 606
 896 ADKDGOMASAPFTSTAIQ-----TYSLSL-VERNENDLAEV--KEENIVKRAAQN 944
 607 TLKRTVTGLAGDRPKDFHFEIELKNNKOELLSTQV-----KT 643
 945 AVAGVTYALDGS-TNKFVF-----TPSQELKAGTVSVYIDGVRDKVNTISKYTSFT 998
 644 DKTN-----LEPDGKATINLKHGESLTQ--GLPEGYSYLVKETSDESGYK--YKYN 691
 999 VSNAPTLSSISADG--AVNVDSKRTITLESDSVNP-TITLKADGTSFTNYTLVNVN 1055

OY 692 SOEVANATVSKTGITSDETLAFE--NNKEPVPTGVODK---INGYIAL--IVIGISLG 744
 DB 1056 NENNTYKIVFHKVGTLDDEFOTYELAVSKDFOTGIDIDSKVFTITGSAVTDVYKALVSGV 1115
 OY 745 IW 746
 DB 1116 SW 1117
 RESULT 9
 SACS3_YEAST STANDARD: PRT: 1301 AA.
 ID SACS3_YEAST STANDARD: PRT: 1301 AA.
 AC P46674;
 DT 01-OCT-1995 (Rel. 32, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Leucine permease transcriptional regulator.
 GN SACS3 OR LEPI OR YDR159W OR YDR358.13.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97027306; PubMed=8873450;
 RA Bauer A., Koelling R.;
 RT "Characterization of the SACS3 gene of Saccharomyces cerevisiae.";
 RL Yeast 12:965-975(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288C / AB972;
 RA Murphy L., Richards C., Harris D., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 189-1301 FROM N.A.
 RX STRAIN=DEY939;
 RA Stella C.A., Korch C., Ramos E.H., Mattoon J.R.;
 RT "Cloning and sequencing of LEPI1, a gene associated with leucine
 transport.";
 RL Yeast 11:460-460(1995).
 CC -1- FUNCTION: POTENTIAL REGULATOR OF LEUCINE PERMEASE GENE(S)
 CC EXPRESSION. REQUIRED FOR NORMAL MITOSIS.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SACS3 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 247805; CAAB767.1; -;
 DR EMBL; 250046; CA90379.1; -;
 DR EMBL; U35227; AAA79056.1; -;
 DR PIR; S51323; S51323.
 DR SGD; S0002586; SACS3.
 DR GO; GO:0005643; C:nuclear pore; IDA.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0030029; P:actin filament-based process; IGI.
 DR GO; GO:0007067; P:mitosis; IMP.
 DR GO; GO:0006611; P:protein-nucleus export; IGI.
 DR InterPro; IPR005062; SACS3_GANP.
 DR Pfam; PF03399; SACS3_GANP; 1.
 DR Transcription regulation; Nuc: Nuclear protein.
 KW SEQUENCE 1301 AA; 149568 MW; 0679DB1673DDACB CRC64;
 Query Match 3.4%; Score 135; DB 1; Length 1301;
 Best Local Similarity 19.1%; Pred. No. 6.3;
 Matches 139; Conservative 105; Mismatches 307; Indels 176; Gaps 31.

OY 115 PFLGSSSVKKMKYKHKHDISTKPEDYAMSPRTGD---ELNOKLRAVMYNGHP---ONA 167
 DB 611 PFGSGESQQLPQISQSHLST---NPLTFPVHGLDSKQKQOIKRTVTDGSSPFVFDOSA 667
 OY 168 NGIMEGLEPLAIRVTOEAVVYSDNADISPDESEFKRESNVLVTSQSLMQ-ALKQ 226
 DB 668 QNSTVEASKAMHISTTSNGAY---DEKLSSQEEKKRKEEORIEBEKQTLKKQENADKO 724
 OY 227 LIDRLATKMPKQVDDPQLSIFESDEKGDKNKCYOLLGGLVPTPPPGDPMPPN 286
 DB 725 VITQIANDLVKEVNNSSVIVKREFSEAVNRQDFDTWTRLY-----DAFLHER 776
 OY 287 -----OPQTSVLRK-----VAIGDSKLE-----GATLQTFGDN 318
 DB 777 LYLIMDSRALKRNSTLKKKFFEKWQASQAKKRNRLLEKKREITLVSHQLGVPQFK 836
 OY 319 VNSFOARVSSNDIGERIELSDGYTTLTSLNTPAGYSIAEPITFEAGKYVTIIDKQI 378
 DB 837 KSTCLFRTPYKGVNNSFMLSDDKNL--IFSPVDEFNKPATHLTKISKLRPLEMQSI 894
 OY 379 --ENPNKEI---VEPYSVEAYND-----FEERSVLTQNTAKFYAKKNKSSQ 422
 DB 895 YYDMLTKKFPNSLTPANLFTYARDWTSLSNRWLISKENLQTAODSKF---SNNIIS 950
 OY 423 VVYCFNADLKSPPOSED-----GKTMTPDFTTGEVYK-----THI 458
 DB 951 RIICID-DEYEPSDFSDQLLIPNTGYT-NFDIDLEKLLKDDGEELIKLTGISLNTNI 1008
 OY 459 AGRDLEKTYVPRDPTDPFLFKIKV--IEKGYREKQALIEYSGLT-----EQQLR 508
 DB 1009 CFSLLIYWEASAEMLTSESTIKHLKLNRSKNVSVIERIDMLMTFESPHKLEPXL 1068
 OY 509 AATQALITYFTDSAEIDKDKLDYHGFDMDNSTLAVALIVEAODSNPQLT-DLD-- 565
 DB 1069 EISHSYVYKLEKRGYDK-----TLROKRSIAGIHSRSTQLQTKYDOK 1113
 OY 566 --FFIPNNKRYQSILIG---TQWHPEDLVDIRMEKKEVLPV---THN-----LTIR 609
 DB 1114 MKKMLEKKKNKYQOQIGERNYVAHLESHID-ASPRSKRKRPILISTSHSSQFPTPLASR 1172
 OY 610 KTVTGLADRTKDFHEIELKNNKQELLSTQVTKDTKLTLEKRD-----GKATINLKHG 662
 DB 1173 LNTGSSSTPPLPSHLAKFKRKNRVTSLHTVLPVSTPSHNNIPAFSGNNNTDI--- 1229
 OY 663 ESLTLQGLPEGYSYLVKRTDSEGYKVKVNSDEVANATYSKTGITSDELAEENKEPVVP 722
 DB 1230 -----QSQQLIENCKSTVYLVNNYSERILGNOEICQPI-----NPVTP 1268
 OY 723 T--GVDQ 727
 DB 1269 VLDGADQ 1275
 RESULT 10
 SLPH_BRECH STANDARD: PRT: 1116 AA.
 ID SLPH_BRECH STANDARD: PRT: 1116 AA.
 AC P38538;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Surface layer protein precursor (Hexagonal wall protein) (HWP).
 OS Brevibacillus choshuensis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Brevibacillus.
 OX NCBI_TaxID=54911;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 54-68.
 RX STRAIN=HPD31;
 RX MEDLINE=90170842; PubMed=2307650;
 RA Ebisu S., Tsuboi A., Takagi H., Naruse Y., Yamagata H., Tsukagoshi N.,
 RA Uda S.;
 RT "Conserved structures of cell wall protein genes among protein-
 RT producing Bacillus brevis strains.";
 RL J. Bacteriol. 172:1312-1320(1990).

```

CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC HEXAGONAL S-LAYER.
CC -1- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90050; BAA14103.1; -
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH; 2.
DR PROSITE: PS01072; SLH DOMAIN; 2.
KM Signal: Cell wall; S-layer; Repeat.
FT SIGNAL: 1 53
FT CHAIN: 54 1116 SURFACE LAYER PROTEIN.
FT DOMAIN: 57 120 SLH 1.
FT DOMAIN: 121 171 SLH 2.
FT DOMAIN: 172 231 SLH 3.
SQ SEQUENCE 1116 AA; 123397 MW; 86D583D7AC72546F CRC64;

Query Match 3.4%; Score 132.5; DB 1; Length 1116;
Best Local Similarity 18.9%; Pred. No. 7;
Matches 162; Conservative 112; Mismatches 266; Indels 315; Gaps 45;

QY 53 LVESSTPMAINPDSSEY-----RWGYESTYVG-----HPY 84
DB 143 IVKGFPPKSFKNQNTVAEAVTMIVRAGLGEPSVAGVSPNMSKSGELINAKGINPW 202
QY 85 YKQFRVA-----HDLRVNLEGSRYQYCFNLKRAFPGLSD-----SSYKKWKKH 130
DB 203 MQQFATIRKMDNALRVKLMEOIE-----GTDRLAVTDETLITKLIK-- 247
QY 131 DGISTKEEDYAMSPRTIGDEL-----NOKRLAVMYNGHPQONANG-----IMEG 173
DB 248 --VTYVDMQMAHEKGNNSDELPLYTNVPAIGLSLKANEVTLNGKADLGSNTTYKVAEG 305
QY 174 LEPINLRVYQGEAVWYYSNAPISNDESEFKRESENLVSTQSLMRQAKLIDPNLA 233
DB 306 IN-NAPDQKQVWMLKIDRENTI---VMEGSEDEEDVMDVSAIYLRGKAFETD-DIV 359
QY 234 TKMPKQVPPDFOLSTFSEDEKGDKNYKGYONLLSGGLVPTKPTPDGPPMPRQPTTSV 293
DB 360 KDLKSDLDLVKLEMGS----- 378
QY 294 LIRKVAIGDYSKLLEGATLQLT--GDNVNSFOARFSSND-IGERIELSDGYTL----- 345
DB 379 --KSYRLTEDTKI---TYNFRFNDPVDAI--SKIKYNDTEFGVKVVLNDNNEVAYLHI 431
QY 346 ---TELNSPAGSIAPIITFVEA-----GKYVTI-IDGKOIEN 380
DB 432 DQGITDKSKYKGYKSVISKIDADKKTKTNLDNSKFSLEDDODESKDPLVLDGPAKY 491
QY 381 PKKEIYPSV-EAYNDFEEFVLTQNYAKFYAKNKGSSQVYVCFNADLKSPDSESD 439
DB 492 GDLKESDYVYVYADGDKDLYLFANRNAE-----GKVEKVSRN---KTDRLVY 540
QY 440 GKRT--MTPDFTTGEEKYTHIAGRLFKTT---VKPPD-----TPPDFLAKIK-- 483
DB 541 GKRTYKVPD-----ASYSENANKDKYKNSDLISLNDGEEVKLLPLPSGRVRIETK 595
QY 484 -----KVIKGG--YREKGAILEYSGLTETQLAATQOLAIFYTDSAEIDKDKLKY 532
DB 596 DAIDRRKPLALITKGFYNSKOTDFTVMTQ---KGRTOI-----VSLDQKIYDR 644
QY 533 HG--FGDMNDSTLAVKILVEYQA-----DSNPQLDLDL-----FIPN 570
DB 645 YGVNVDKSKNDKQAFKDLVELLQPKYVKEDSATDANQVLELVNDSNGEVDKQYVLD 704
QY 571 NKKYOSLIGTQW-----HPEDLVDIRMEDKEVIVPTIHNLTLRKTVTGLAGDRTKDHF 625

```

```

DB 705 KLYSE--KSTWMDKLADEDDVDVGYEDVDKTAVFMMTGDLT--PATG-----TK----- 750
QY 626 EELKN-----NKDEL-LSQTVKNDKTLF----- 649
DB 751 RGLKNAAGTAKFKVDKAKSKDLKWKVSYVDEKGEVQALFYVDGSLGGLDHPGKWKQYGYA 810
QY 650 FKDKKATITNLKHGESITL-----LQGIPEGYSYLKKEETDESEGYKV--KVNQOEVAANAFVSK 702
DB 811 SKQDITITVTKQSDSYTEKEKYKLDG-----DADDLKVQDDIRGRVDISTLNS 858
QY 703 TG-TTSDETLAEFNN 716
DB 859 DGEVIVDDVEEVVNN 873

RESULT 11
CADN_DROME
ID CADN_DROME STANDARD; PRT; 3097 AA.
AC 015943; Q9VJB7;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 41, Last annotation update)
DE 28-FEB-2003 (Rel. 41, Last sequence update)
DE Neural-cadherin precursor (Cadherin-N protein) (DN-cadherin).
GN CADN OR CG7100.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phnydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM D).
RC TISSUE=Embryo, and Head;
RX MEDLINE=97388431; PubMed=9247265;
RA Iwai Y., Usui T., Hirano S., Steward R., Takeichi M., Uemura T.;
RT "Axon patterning requires DN-cadherin, a novel neuronal adhesion
RL receptor, in the Drosophila embryonic CNS. ";
RN Neuron 19:77-89(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agdayani A., An H.-J., Andrews-Plankkoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burks C.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin R.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegum G.,
RA Jalali M., Kalush E., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,
RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

```

```

RA Wang Z., Massarini D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Wooley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195(2000).
RL [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RC STRAIN-Berkeley;
RC MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Klinkner J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu U., Betman B.P.,
RA Beltenkova B.R., Gelinker S.E., de Grey A.D.N.J., Dysdale R.A.,
RA Harris N.L., Richter J., Russel S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review".
RT Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RL [4]
RN INTERACTION WITH ARM.
RP MEDLINE=98298928; PubMed=9635189;
RA Loureiro J., Pelfer M.;
RA "Roles of Armadillo, a Drosophila catenin, during central nervous
RT system development.";
RT Curr. Biol. 8:622-632(1998).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. MAY ASSOCIATE WITH ARM NEURAL
CC ISOFORM AND PARTICIPATE IN THE TRANSMISSION OF DEVELOPMENTAL
CC INFORMATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing: Named isoforms=8;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=D;
CC IsoId=O15943-1; Sequence=Displayed;
CC Name=A;
CC IsoId=O15943-2; Sequence=VSP_000667, VSP_000668;
CC Name=B;
CC IsoId=O15943-3; Sequence=VSP_000668;
CC Name=C;
CC IsoId=O15943-4; Sequence=VSP_000667, VSP_000668, VSP_000669;
CC Name=E;
CC IsoId=O15943-5; Sequence=VSP_000667;
CC Name=F;
CC IsoId=O15943-6; Sequence=VSP_000669;
CC Name=G;
CC IsoId=O15943-7; Sequence=VSP_000667, VSP_000669;
CC Name=H;
CC IsoId=O15943-8; Sequence=VSP_000668, VSP_000669;
CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, THE PROTEIN FIRST APPEARS IN
CC THE MESODERM AT STAGE 9 AND IS PRESENT IN THE MYOBLASTS AND MUSCLE
CC FIBERS BY STAGE 12 AND STAGE 14, RESPECTIVELY. AT STAGE 12 THE
CC PROTEIN IS ALSO LOCATED IN THE AXONS OF THE ENTIRE CNS, BUT NOT IN
CC THE GLIAL CELLS. IN THIRD INSTAR LARVAE PROTEIN IS EXPRESSED IN
CC THE CNS NEUROPILE, PHOTORECEPTOR AXONS AND PRECURSORS OF ADULT
CC MUSCLES.
CC -1- SIMILARITY: Contains 16 cadherin domains.
CC -1- SIMILARITY: Contains 3 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

ID	Gene	Protein	Function
CC	EMBL; AB002397;	BAA22131.1;	-
DR	EMBL; AE003656;	AAF53635.1;	-
DR	EMBL; AE003656;	AAN10992.1;	-
DR	EMBL; AE003656;	AAN10993.1;	-
DR	EMBL; AE003656;	AAN10994.1;	-
DR	EMBL; AE003656;	AAN10995.1;	-
DR	EMBL; AE003656;	AAN10996.1;	-
DR	EMBL; AE003656;	AAN10997.1;	-
DR	PIR; T00021;	T00021.	-
DR	HSSP; P00740;	IBDM.	-
DR	FlyBase; FBgn005911;	C:intercellular junction; IDA.	-
DR	GO; GO:0008014;	F:calcium-dependent cell adhesion molecule ac. .; IPI	-
DR	GO; GO:0007412;	P:homon target recognition; IMP.	-
DR	GO; GO:0007156;	P:homonophilic cell adhesion; IDA.	-
DR	InterPro; IPRO02126;	Cadherin.	-
DR	InterPro; IPRO00233;	Cadherin_C_term.	-
DR	InterPro; IPRO00742;	EGF_2.	-
DR	InterPro; IPRO01881;	EGF_Ca.	-
DR	InterPro; IPRO06209;	EGF_Like.	-
DR	InterPro; IPRO01791;	Laminin_G.	-
DR	Pfam; PF000028;	cadherin; 14.	-
DR	Pfam; PF01049;	Cadherin_C_term; 1.	-
DR	Pfam; PF00008;	EGF; 3.	-
DR	Pfam; PF00054;	Laminin_G; 2.	-
DR	PRINTS; PRO0205;	CADHERIN.	-
DR	SMART; SM00112;	CA; 16.	-
DR	SMART; SM00179;	EGF_CA; 1.	-
DR	SMART; SM00282;	Lamc; 2.	-
DR	PROSITE; PS00232;	CADHERIN_1; 9.	-
DR	PROSITE; PS50268;	CADHERIN_2; 16.	-
DR	PROSITE; PS00022;	EGF_1; 3.	-
DR	PROSITE; PS01186;	EGF_2; 3.	-
DR	PROSITE; PS50025;	LAM_G_DOMAIN; 2.	-
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;		-
KV	Signal; EGF-like domain; Alternative splicing.		-
FT	SIGNAL	1	36
FT	PROPEP	37	?
FT	CHAIN	?	?
FT	DOMAIN	? 1454	NEURAL-CADHERIN.
FT	TRANSSEM	1455 1475	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	1476 3097	POTENTIAL.
FT	DOMAIN	181	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	430 543	CADHERIN 1.
FT	DOMAIN	554 651	CADHERIN 2.
FT	DOMAIN	660 756	CADHERIN 3.
FT	DOMAIN	766 858	CADHERIN 4.
FT	DOMAIN	867 968	CADHERIN 5.
FT	DOMAIN	978 1078	CADHERIN 6.
FT	DOMAIN	1087 1183	CADHERIN 7.
FT	DOMAIN	1193 1299	CADHERIN 8.
FT	DOMAIN	1307 1414	CADHERIN 9.
FT	DOMAIN	1423 1514	CADHERIN 10.
FT	DOMAIN	1523 1630	CADHERIN 11.
FT	DOMAIN	1639 1742	CADHERIN 12.
FT	DOMAIN	1749 1861	CADHERIN 13.
FT	DOMAIN	1870 1966	CADHERIN 14.
FT	DOMAIN	1974 2085	CADHERIN 15.
FT	DOMAIN	2346 2377	CADHERIN 16.
FT	DOMAIN	2379 2585	EGF-LIKE 1.
FT	DOMAIN	2592 2627	LAMININ G-LIKE 1.
FT	DOMAIN	2631 2822	EGF-LIKE 2.
FT	DOMAIN	2869 2902	LAMININ G-LIKE 2.
FT	DISULFID	2346 2357	EGF-LIKE 3.
FT	DISULFID	2351 2366	POTENTIAL.
FT	DISULFID	2368 2377	POTENTIAL.
FT	DISULFID	2592 2607	POTENTIAL.
FT	DISULFID	2601 2616	POTENTIAL.
FT	DISULFID	2618 2627	POTENTIAL.
FT	DISULFID	2869 2880	POTENTIAL.

Query Match 3.4%; Score 132.5; DB 1; Length 3097;
 Best Local Similarity 19.8%; Pred. No. 32;
 Matches 114; Conservative 89; Mismatches 192; Indels 181; Gaps 29;

QY 287 QPOTSVLIRKIA-IGDYSLKLGALQLT--GDVNSFQARFSSNDIGRIELSDGT 343
 DB 719 QPHTVSLIVATEDSGFSTSV-D-LTRIVDVNDMAKPELPDQAQANVEDIPLGTSIL 777
 QY 344 TLELNSPAG-----YSIAEPIPFKEACKVYTIIDGKQIENKRIEVPYSVEAVND 397
 DB 778 RVAMSDSGSNAIEIYVSD--HFAVDSNGI--YANKQOLDNANNAYEFITAKDKG 834
 QY 398 E-EFSLTQYNAKFEYAKNKS-----SQVYCFENADLKSP-----DSED 439
 DB 835 EPPKSGVAVR-----VYTKNKNDDEPKFSQYVTPVDNAGPNTLVTVASDKGDN 889
 QY 440 -----GKTMTPDFT-----TGEVKYTHA-----GRDLFKTYV----- 468
 DB 890 VREGVGGTSGSQFYIEDITGVIRLHNKAISLDKREYELNVAMDDSCCVNGDQTIHT 949
 QY 469 -----KPRDTPDPTFLKHK-----KVI-----EKGYREKQAIEX 499
 DB 950 STAVVVFITDVNDKRVFVDCSTYRKVEGAPNGSPVILKVAATDEKGV--NGQ-VKY 1006
 QY 500 SGLTEQLRAATQALAIYFTDSAEIDKDKLKDYG-----FGDMNDSTL-AVAK 547
 DB 1007 S-IVQGPNOGKFTV--DETEGEVSTNKVFEDREGDGKFTVSVYKATDGDPSLEGVCS 1063
 QY 548 ILVEYADSNRPQTLDTDFIPNNKYSQSLIGTOMHEDLVDIRMDEKKEVLPVTHNL 607
 DB 1064 FTVETIDVNDNPPLFPRQKV--ENVKODASIGT-----NLIRVS----- 1101
 QY 608 LRKTVGLADGRTKDFEIELEKNNKQELLSQTVKTDKTNLEFKDGATINLKHSBTL 667
 DB 1102 -----ASDEAD-----NNGAIYSLTAPENPDLFEPIQA----- 1133
 QY 668 QGLEPGSYLVKETDSEGVYKVNSEVANAVYSGTISDETLAENKPEVPTGVD 727
 DB 1134 ---ESGMIVLKKPLDRETYKLEAMADKGYPLSRFEVQIDVVDRAVN-----PPWMDH 1185
 QY 728 KINGYLAL-----IYAGISLIGWGHITRIR 754
 DB 1186 TVYGPITYVENMPGVGKVASIKASSGLEGNPTFYR 1221

RESULT 12
 LHS1_YEAST STANDARD; PRT; 881 AA.
 AC P36016;
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Heat shock protein 70 homolog LHS1 precursor.
 GN LHS1 OR YKJ073W OR YKJ355.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pohl T.M., Pohl F.M.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=5288C;
 RX MEDLINE=94378724; PubMed=8091863;
 RA Rasmussen S.W.;
 RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
 RT NUP100 gene, an open reading frame (ORF) possibly representing a
 RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
 RT addition to seven ORFs with weak or no significant similarity to
 RT known proteins";
 RL Yeast 10:569-574(1994).

RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=96256281; PubMed=8654361;
 RA Craven R.A., Egerton M., Stirling C.J.;
 RT "A novel Hsp70 of the yeast *Er* lumen is required for the efficient
 RT translocation of a number of protein precursors";
 RL EMO J. 15:2640-2650(1996).
 CC -1- FUNCTION: PROBABLY ACTS AS A CHAPERONE INVOLVED IN BOTH
 CC POLYPEPTIDE TRANSLOCATION AND SUBSEQUENT FOLDING.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC -1- PTM: N-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X75780; CA53401.1; -;
 DR EMBL: Z28073; CA81910.1; -;
 DR PIR: S37895; S37895.
 DR SGD: S0001556; LHS1.
 DR GO: GO:0005783; C: endoplasmic reticulum; IDA.
 DR GO: GO:0003754; F: chaperone activity; IMP.
 DR GO: GO:0015031; P: protein transport; ICI.
 DR GO: GO:0006986; P: response to unfolded protein; ICI.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR01023; Hsp70.
 DR Pfam: PF00012; Hsp70; 1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR ProDom: PD000089; Hsp70; 1.
 DR PROSITE: PS00014; ER_TARGET; 1.
 DR PROSITE: PS00329; HSP70_2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 DR PROSITE: PS00297; HSP70_1; FALSE NEG.
 KW Chaperone; Endoplasmic reticulum; Glycoprotein; signal; Arg-binding.
 FT STGNAL 1 20
 FT CHAIN 21 881
 FT CARBOHYD 128 128
 FT CARBOHYD 458 458
 FT CARBOHYD 474 474
 FT CARBOHYD 481 481
 FT CARBOHYD 489 489
 FT CARBOHYD 527 527
 FT CARBOHYD 844 844
 FT SITE 878 881
 SQ SEQUENCE 881 AA; 99571 MW; ACED092CA3A34785 CRC64;

Query Match 3.3%; Score 130.5; DB 1; Length 881;
 Best Local Similarity 18.3%; Pred. No. 6.4; Indels 199; Gaps 32;
 Matches 132; Conservative 119; Mismatches 199;

QY 52 GLVESSTPNAINDSSS-----EYRWYGEYSYRG-----HPYKOF 88
 DB 241 GSIAKAMFSLQPEDTTPVTEFEGYVPHGAKFTMDISLKNFLEHNPALR-- 298
 QY 89 RVAHDLRVNLEGRSRQVYCFNLKKAEPJLGSDDSVKWKYKHKDICTKEDYAMSRTIG 148
 DB 299 --TDELHANPKALAKINQAEKAKLILSANSEASI-----NIESLINDIDFRTSTR 348
 QY 149 DELNOKLRVAMVNGHQNNANGIMELEPNAIRVTOEAVWYSDNAPISNPDSFKREE 208
 DB 349 QEFEEFT-----ADSLIDIVKPLIND-AVTKQ-PGGGTNP-----E 383
 QY 209 SNLV-----STQSLNRQALQQLIDPNLTKMKQVDDFOLSIPESEDKGDKYKNGYON 264
 DB 384 INVILLAGSSRRPIYQDLIKVSE-----KVLARNNAD-ESAVNGVWGRGKILNSFK- 438
 QY 265 LLSGLVPTKPPPTGPPMP-----PNOPTTSLIRKVAIGDYSLKLGATLQJGN 318

Db 439 -----TKPLNVDSVNTYSEKLSNESELYDFTRGSAVYPKNTSILTNT-----DS 485
 Qy 319 V-NSFOARVSSNDIGERIELSDGTITTELNSPAGYSIAPRIFPKVAGVYITID--- 374
 Db 486 IPNNFTIDLFENGKLFETITVNSAINKNSYSSDKSCSSGVANITFDLSDBRFSIQEVNC 545
 Qy 375 -----GKQIENPNK-----EIVEPYSEVAYNPFEEFSVLTQONAKFYAK 415
 Db 546 ICQSENDIGNSKQIKNNKSRLATFSEDEVEIKRLSPSESRLEHKKLIDKDKKEFQOE 605
 Qy 416 NKNGSSOVVYCFNADLSPDESGKTMPTDFTGVEKYTHIAGRDLEKTYVAPRDTDP 475
 Db 606 NLNVLESNTLYARNLIMLMDVEVQNGPKSQVEELS--EMWVYLLDMLJEDASF-----DTPD 658
 Qy 476 DTFPKHT-----KKVIE-----KGYRKG-----QAIYSGLTETQLAA 510
 Db 659 EDIVSRIREIGILKKKLELYNDSAKEPINSOQFGMEEGHKLLOAITHRNVEEFLSQ 718
 Qy 511 TOLAIYFTDSAEIDKDKLDYHGFQDMNDSTLAVAKILVEYAO--DSNPOLFDLPFI 568
 Db 719 FETEFADTIDNVREFFKKI-----QPAYSKALSTWEEITLTSFKNSISIEKFL 768
 Qy 569 PNNKYSGLGTFQHPEDVDI-----IMEDKKEVIVTHNLTKVTYGLAGDRT 620
 Db 769 AKN-----LFGEDLR-EHLFEIKLOFDWYRTKLEKRLIK-----SGDES 808
 Qy 621 -----KDPHF-EELKNNKQELLQSV-----KIDKTNLEKCKOKAT-NLKHEES 664
 Db 809 RLNEIKLHLRNFRLQKREKRLKLEQKSRNNNETESTVINSADKTTIYNDKTES 868

RESULT 13

DPOL_BORBU STANDARD: PRT: 908 AA.
 ID DPOL_BORBU

AC 051498;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA polymerase I (EC 2.7.7.7) (POL I).
 GN POLA OR BB0548.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxId=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Caajens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kellavag A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uterback T., Wathey L., McDonald L., Artlach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi."
 RL Nature 350:580-586(1997).
 CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate + (DNA)(N).
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC

DR EMBL: AE001156; AAC66909.1; -
 DR PIR: C70168; C70168.
 DR HSSP: P19821; IKTQ.
 DR TIGR: BB0548; -
 DR InterPro: IPR002562; 3_5_exonuclease.
 DR InterPro: IPR002421; 5_3_exonuclease.
 DR InterPro: IPR001098; DNA_POL.
 DR InterPro: IPR002298; DNA_POL.
 DR InterPro: IPR000513; Exo_N.I.
 DR InterPro: IPR003584; HNH_2.
 DR Pfam: PF01612; 3_5_exonuclease; 1.
 DR Pfam: PF02739; 5_3_exonuclease; 1.
 DR Pfam: PF01367; 5_3_exonuclease; 1.
 DR Pfam: PF00476; DNA_POL_A; 1.
 DR PRINTS: PR00686; DNAPOL1.
 DR SMART: SM00474; 3EXOC; 1.
 DR SMART: SM00475; 53EXOC; 1.
 DR SMART: SM00279; HNH2; 1.
 DR SMART: SM00482; POLAC; 1.
 DR TIGRFAMS: TIGR00593; pola; 1.
 DR PROSITE: PS00447; DNA_POLYMERASE_A; 1.
 DR Translase; DNA-directed DNA polymerase; DNA replication; DNA repair;
 KM Hydrolase; Exonuclease; DNA-binding; Complete proteome.
 FT DOMAIN 1 318
 FT DOMAIN 319 531
 FT DOMAIN 532 908
 SQ SEQUENCE 908 AA; 105503 MW; B58512AE0DCBD94 CRC64;

Query Match

Best Local Similarity 19.3%; Score 130.5; DB 1; Length 908;
 Matches 166; Conservative 103; Mismatches 264; Indels 327; Gaps 43;

Qy 13 LNTQVLSKNSKRFVTVLGVFLMFLVTSWVGKTYFGLVESSTPAINPDSSEFRW 72
 Db 26 LNTQ-----GENVNAF-----IGFF-----KTLFPIKEKNPRLITTFDESVPT 65
 Qy 73 YGESYVAGHPYKQFR-----VANDLRVNL--- 98
 Db 66 FRKQKTY-----PSYKATROLPPDDLIPQIGWIKKGLAKKPIFEMEGETEADLASFRKK 121
 Qy 99 EGSRSYQVYCFN-----LKKAPPLGSSSVK---KWKKHGDI-STKPEDYAMS 143
 Db 122 AAKNNVLYIISPOKDLQTMSEYVKILKIENNSFIENDNEVYTKKFGVNSQIOLDYL- 179
 Qy 144 PRITGDELNOKLRVAVMNGHPOMAN-----GIMGLEPLN-----AIRVQEVW 188
 Db 180 -AIVGDR-SDNIPGKIGIGAKGAANLREFTLDTGIVSNLETKKHRELLIKENAF 237
 Qy 189 YV-----SDNAPISMPD-----ESFKRESESNVSTQSLMGMQALKOLIDPRLATKMKQ 239
 Db 238 STELVSELENKIPETIENPALKNESEELISLPEKHSATALLIKYKKDL-----KQKE 291
 Qy 240 VPDPEOLSTFSESDGDKRYNKGQVNLGSLGVTPKPTPGDPMPBPNOPTTVLIRKYA 299
 Db 292 NAD--QKSLPFQGE-----PTNSL----- 308
 Qy 300 IGVSKLLEGATLQLTGNNVSPQARVSSNDIGERIELSDGTITL-----TELNS 350
 Db 309 -----DDIMTIDTENVKYRSITFKIELDLIELSKRAKYSIDTFERS 351
 Qy 351 PAGVS---IAEPIPEK-----VEA-GKYVTIIDGQIENPNKEIPEPSVEAYND-F 397
 Db 352 LDYTAKLIGISISKKEGEGYIPIEANGKIT-----IEKNYTIQFNMLF 397
 Qy 398 EEFVLTQNTYAKFYAAKNNKSSQVYVCFNADLSPDSESGKTMPTDF---TGGEK 454
 Db 398 ESNPRLIQNV-KPDYKILKNNG-----FN--PIPPYFD---TMAAVLIDNSVYS 443
 Qy 455 YTHIAGRLPFTYVAPRPTDPTFLKHKKYIEKYEKRGGAIEVSGLTQLRATOLA 514
 Db 444 LDPLAEKYLIMKNIKYED-----VIQKNDNFANISLEWATSYSSBDADITPRL 491
 Qy 515 IYFTDSALDK-DKLKDYHGFQDMNDSTLAVAKILVEYAO----- 555

DB 492 FNIFTKKEDKIDLK-----MHEIEMFNKVIIEEMENCIYLDKEYLKEGRELGR 543
 OY 556 -----SNPOLND-----LDFIPNNKKQSLIGTQWHEEDLVDI 591
 DB 544 ELEAIENEIETKSGIDFNPNSPKOMHEILFEKLNKLPEKMK-----KSDTDIK 592
 OY 592 RMEDEKKEVLPVHTNLTLRKTVTGLADRRKDFHFEIEELKNNKOELLQGVTKDTNLEFK 651
 DB 593 VESLEQREHSEIENLIKRYQIAKLKSTYTDNL--LELN-----YKTNRLHSTFI 640
 OY 652 DQKATINLKHGSELTIOGPEGSIYKENDSGYV-KVNSEVNAVATSK-----702
 DB 641 QFTATGRTSTINPNQNTP-----IK--DEKGRIRKRAFKPENGNIFISADYSQIEIA 692
 OY 703 --TGTSDETL--AFENKKE 718
 DB 693 ILAHLSDDEVILKAFENKMD 712

RESULT 14
 RPGR_HUMAN
 ID AC 092834; 000702; 000737; 093039; Q9HD29; Q9UMK1;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE X-linked retinitis pigmentosa GTPase regulator.
 GN RPGR OR RP3 OR XLRP3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4), AND VARIANTS RP3 GLN-98;
 RP VAL-215; ARG-250 AND 296-THR--ILE-300 DEL.
 RX MEDLINE=96241570; PubMed=8673101;
 RA Meindl A., Dry K.L., Herrmann K., Manson F.D., Ciccodicola A.,
 RA Edgar A.J., Carvalho M.R.S., Achatz H., Heilebrand H., Lennon A.A.,
 RA Migliaccio C., Porter K., Zrenner E., Bird A.C., Jay M., Lorenz B.,
 RA Wiltner B., D'Urso M., Meilinger T., Wright A.;
 RT "A gene (RPGR) with homology to the RGC1 guanine nucleotide exchange
 RT factor is mutated in X-linked retinitis pigmentosa (RP3).";
 RL Nat. Genet. 13:35-42(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS RP3 CYS-130; SER-235 AND
 RP SER-275.
 RC TISSUE=Retina;
 RX MEDLINE=96414315; PubMed=8817343;
 RA Roepman R., van Duijnhoven G., Rosenberg T., Pinckers A.J.L.G.,
 RA Bleeker-Wagemakers L.M., Bergen A.A.B., Post J., Beck A.,
 RA Reimhardt R., Ropers H.-H., Cremers F., Berger W.;
 RT "Positional cloning of the gene for X-linked retinitis pigmentosa 3:
 RT homology with the guanine-nucleotide-exchange factor RGC1.";
 RL Hum. Mol. Genet. 5:1035-1041(1996).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC TISSUE=Retina;
 RX MEDLINE=99330567; PubMed=10401007;
 RA Kirschner R., Rosenberg T., Schultz-Hendbrok R., Lenzner S., Feil S.,
 RA Roepman R., Cremers F.P.M., Ropers H.-H., Berger W.;
 RT "RPGR transcription studies in mouse and human tissues reveal a
 RT retina-specific isoform that is disrupted in a patient with X-linked
 RT retinitis pigmentosa.";
 RL Hum. Mol. Genet. 8:1571-1578(1999).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=21558421; PubMed=11702207;
 RA Kirschner R., Erturk D., Zeitz C., Sahin S., Ramser J.,
 RA Cremers F.P.M., Ropers H.-H., Berger W.;
 RT "DNA sequence comparison of human and mouse retinitis pigmentosa
 RT GTPase regulator (RPGR) identifies tissue-specific exons and putative
 RT regulatory elements.";

RL Hum. Genet. 109:271-278(2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Retina;
 RA Berger W.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SOURCE OF 525-840 FROM N.A. (ISOFORM 1), AND VARIANTS RP3 GLN-98 AND
 RP ARG-250.
 RX MEDLINE=20392404; PubMed=10932196;
 RA Verwoort R., Lennon A., Bird A.C., Tulloch B., Axton R., Milano M.G.,
 RA Meindl A., Meilinger T., Ciccodicola A., Wright A.F.;
 RT "Mutational hot spot within a new RPGR exon in X-linked retinitis
 RT pigmentosa.";
 RL Nat. Genet. 25:462-466(2000).
 RN [7]
 RP VARIANTS RP3 VAL-60 AND VAL-75, AND VARIANTS GLY-262; LYS-425; VAL-431
 RP AND GLY-566.
 RX MEDLINE=98153625; PubMed=9399904;
 RA Buraczynska M., Wu W., Fujita R., Buraczynska K., Phelps E.,
 RA Andreasson S., Bennett J., Birch D.G., Fishman G.A., Hoffman D.R.,
 RA Inana G., Jacobson S.G., Musarella M.A., Slevy P.A., Swarcop A.;
 RT "Spectrum of mutations in the RPGR gene that are identified in 20% of
 RT families with X-linked retinitis pigmentosa.";
 RL Am. J. Hum. Genet. 61:1287-1292(1997).
 RN [8]
 RP VARIANT RP3 VAL-60.
 RX MEDLINE=99070804; PubMed=9855162;
 RA Fishman G.A., Grover S., Jacobson S.G., Alexander K.R., Derlacki D.J.,
 RA Wu W., Buraczynska M., Swarcop A.;
 RT "X-linked retinitis pigmentosa in two families with a missense
 RT mutation in the RPGR gene and putative change of glycine to valine at
 RT codon 60.";
 RL Ophthalmology 105:2286-2296(1998).
 RN [9]
 RP VARIANTS RP3 ASN-99 AND VAL-289.
 RX MEDLINE=99415288; PubMed=10482958;
 RA Milano M.G., Testa F., Strazulilo M., Trujillo M., De Bernardo C.,
 RA Grammatico B., Simonelli F., Mangino M., Torrente I., Ruberto G.,
 RA Beneyto M., Antinolo G., Rinaldi E., Danesino C., Venturto V.,
 RA D'Urso M., Ayuso C., Baiget M., Ciccodicola A.;
 RT "Mutation analysis of the RPGR gene reveals novel mutations in south
 RT European patients with X-linked retinitis pigmentosa.";
 RL Eur. J. Hum. Genet. 7:687-694(1999).
 RN [10]
 RP VARIANTS ILE-76; LYS-425 AND GLU-566.
 RX MEDLINE=99408233; PubMed=10480356;
 RA Zito I., Thiselton D.L., Gorin M.B., Stout J.T., Plant C., Bird A.C.,
 RA Bhattacharya S.S., Hardcastle A.J.;
 RT "Identification of novel RPGR (retinitis pigmentosa GTPase regulator)
 RT mutations in a subset of X-linked retinitis pigmentosa families
 RT segregating with the RP3 locus.";
 RL Hum. Genet. 105:57-62(1999).
 RN [11]
 RP VARIANT RP3 ARG-302.
 RX MEDLINE=20202845; PubMed=10737996;
 RA Zito I., Gorin M.B., Plant C., Bird A.C., Bhattacharya S.S.,
 RA Hardcastle A.J.;
 RT "Novel mutations of the RPGR gene in RP3 families.";
 RL Hum. Mutat. 15:386-386(2000).
 RN [12]
 RP VARIANTS LYS-425; GLN-526 DEL; MET-533 AND GLU-566.
 RX MEDLINE=20438359; PubMed=10980513;
 RA Zito I., Morris A., Tyson P., Wnship I., Sharp D., Gilbert D.,
 RA Thiselton D.L., Bhattacharya S.S., Hardcastle A.J.;
 RT "Sequence variation within the RPGR gene: evidence for a founder
 RT complex allele.";
 RL Hum. Mutat. 16:273-274(2000).
 RN [13]
 RP VARIANT RP3 ASP-436.
 RX MEDLINE=21133224; PubMed=11180598;
 RA Guevara-Fujita M., Fahrner S., Buraczynska K., Cook J., Wheaton D.,
 RA Cortes F., Vicencio C., Pena M., Fishman G.A., Mintz-Hittner H.,

